Nucleic Acids Research, 1994, Vol. 22, No. 12 2383–2391

Cloning and functional characterization of LCR-F1: a bZIP transcription factor that activates erythroid-specific, human globin gene expression

John J.Caterina, David Donze, Chiao-Wang Sun, Dominic J.Ciavatta and Tim M.Townes*
Department of Biochemistry and Molecular Genetics, Schools of Medicine and Dentistry, University of Alabama at Birmingham, Birmingham, AL 35294, USA

Received February 2, 1994; Revised and Accepted May 12, 1994 GenBank accession no. U08853

ABSTRACT

DNase I hypersensitive site 2 (HS 2) of the human β-globin Locus Control Region (LCR) directs high level expression of the β-globin gene located 50 kilobases downstream. Experiments in cultured cells and in transgenic mice demonstrate that duplicated AP1-like sites in HS 2 are required for this powerful enhancer activity. A cDNA clone encoding a basic, leucine-zipper protein that binds to these sites was isolated and designated Locus Control Region-Factor 1 (LCR-F1). This protein is a member of a new family of regulatory factors that contain a 63 amino acid 'CNC domain' overlapping the basic region. This domain is approximately 70% identical in the Drosophila Cap N Collar (CNC) protein, NF-E2 and LCR-F1. LCR-F1 transactivates an HS 2/γ-globin reporter gene over 170-fold in transient transfection experiments specifically in erythroid cells. These results suggest that LCR-F1 may be a critical factor involved in LCR-mediated, human globin gene expression.

INTRODUCTION

The human β-globin Locus Control Region (LCR) is a 16 kilobase region of DNA located 6 to 22 kilobases upstream of the ε-globin gene (Figure 1). Five DNase I hypersensitive sites have been mapped in this area and four of these sites (5′ HS 1–4) are erythroid-specific and developmentally stable; that is, they are present in erythroid cells at all stages of development (1–3). The functional significance of these sequences has been demonstrated in numerous experiments in which the sites are linked to ε-, γ- and/or β-globin genes and expression is assayed in transgenic mice or in cultured erythroid cells [for reviews see (4–7)]. In transgenic mice 5′ HS 1–5 sites enhance the expression of a linked β-globin gene 300-fold (8, 9). Individual sites 5′ HS 2, 5′ HS 3 and 5′ HS 4 activate expression approximately 100 fold in transgenic animals and 5′ HS 1 lacks significant activity (9–13); the 5′ HS 5 site may contain a domain boundry sequence that insulates globin gene family members from effects of upstream regions (14). The importance of sites 5′ HS 2–4 has been confirmed by deletion of these sequences in an Hispanic β-thalassemia patient. This deletion results in the absence of ε-, γ-, and β-globin expression from the mutant chromosome even though all gene family members are intact (15). Furthermore, Forrester *et al.* (16) have demonstrated that the entire β-globin locus in this patient, including a region of DNA located 100 kb downstream of the β-globin gene, is resistant to DNase I digestion in erythroid cells. These results suggest that the LCR has 2 important activities. First, the sequences 'open' a 200 kb β-globin chromosomal domain specifically in erythroid cells and secondly, they serve as a master enhancer for ε-, γ-, and β-globin gene expression.

To define the sequences responsible for these activities we and others footprinted nuclear proteins from erythroid and non-erythroid cells on DNA fragments containing 5′ HS 2. Individual factor binding sites were then deleted and the mutant HS 2 β-globin constructs were tested in erythroid culture cells (17–19) or in transgenic mice (20–22). Deletion of duplicated AP1-like sites in 5′ HS 2 virtually eliminated enhancer activity. These results demonstrated the functional significance of the AP1-like sites and suggested that proteins which bind to the sites are critical for human β-globin gene expression.

Two functional assays suggest that AP1 is not responsible for the activity of these sites. First, overexpression of c-Jun in the human erythroleukemia cell line K562 does not activate but significantly inhibits globin gene expression (18). Secondly, a G to T transversion located 2 bp upstream of the AP1 consensus sequence in 5′ HS 2 (Figure 1) does not inhibit AP1 binding in vitro but severely inhibits expression of HS 2 β-globin gene constructs in erythroid cells (17–19). These data suggest that proteins other than AP1 bind to the AP1-like sites in 5′ HS 2 and activate high levels of globin gene expression.

MATERIALS AND METHODS

Transgenic animal experiments

The 5′ HS2 (K-P) β fragments were purified from plasmid sequences and analyzed in mice as described (21). Human and mouse β-globin levels in 16-day fetal livers were determined by

*To whom correspondence should be addressed

primer extension and by solution hybridization with human and mouse β-globin oligonucleotides (9, 23). The fraction of expressors represents the number of transgenic animals that express the transgene. Expression/gene copy is the mean percent expression per gene copy; this value is calculated as follows: (human β mRNA/human β gene copy)/(mouse β mRNA/ mouse β gene copy) ×100. Site-specific mutations in 5′ HS 2 (K-P) β were made in the pSELECT system as described by the manufacturer (Promega). The mutagenic oligonucleotides were as follows:

G>T 5′ CAAGCACAGCAATTCTGAGTCATTATGAGTCATGCTGAGGC 3′
A>C 5′ CAAGCACAGCAATGCTGCGTCATGATGCGTCATGCTGAGGC 3′

Construction and screening of K562 cDNA library

Poly A RNA was purified from uninduced K562 cells and double-stranded cDNA was prepared with the Superscript cDNA kit from Bethesda Research Labs (BRL) as described by the supplier. EcoRI−NotI adaptors (BRL) were ligated onto the termini. cDNAs larger than 500 bps were purified on a Sephacryl S-500 HR column (BRL) and ligated with EcoRI digested, dephosphorylated λgt11 arms (BRL). DNA was packaged (Packagene, Promega) and 5 × 10^5 phage were plated onto Y1090 cells at 20,000 phage/15 cm plate. The plaques were blotted onto nitrocellulose and probed with multimerized 5′ HS2 AP1-like sites as described (24, 25). The double stranded oligonucleotides were as follows:

AP1 5′ ATGCTGAGTCATGATGAGTCATGCTGAGGCTTAGGG 3′
3′ ACTCAGTACTACTCAGTACGACTCCGAATCCCTACG 5′
G>T 5′ ATTCTGAGTCATTATGAGTCATGCTGAGGCTTAGGG 3′
3′ ACTCAGTAATACTCAGTACGACTCCGAACCCTAAG 5′
A>C 5′ ATGCTGCGTCATGATGCGTCATGCTGAGGCTTAGGG 3′
3′ ACGCAGTACTACGCAGTACGACTCCGAATCCCTACG 5′
C/EBP 5′ AATTCAATTGGGCAATCAGG 3′
3′ GTTAACCCGTTAGTCCTTAA 5′.

The 4 bp sequences at the 5′ ends of the oligonucleotides are complementary to allow the formation of head to tail concatomers. The two strands were annealed and the double stranded oligonucleotides were ligated to form multimers. Multimers of 150−300 base pairs were purified on agarose gels and labeled by nick translation (26). Following the isolation and characterization of the initial clone, a second K562 cDNA library (Clontech) was screened with the 1.7 kb LCR-F1 insert. Clones that overlapped the 5′ and 3′ ends of the original clone were isolated and sequenced.

Northern blot hybridizations

Five or ten micrograms of total RNA from cell lines or primary tissues were fractionated on a 1.5% agarose gel containing 1.85% formaldehyde. RNA was blotted to a Duralose Membrane (Stratagene), UV crosslinked in a Stratagene Stratalinker (autosetting) and baked *in vacuo* at 80°C for 30 min. The blots were probed with a randomly labeled, 0.96 kb BamHI−EcoRI fragment containing the 3′ end of the LCR-F1 cDNA.

Gel shift analysis

Gel shift eperiments were performed as described (27, 28). Briefly, 20,000 cpm of 5′ end-labeled oligonucleotides containing the duplicated AP1-like sites from 5′ HS 2 were mixed with a 100-fold molar excess of the indicated, unlabeled oligonucleotides and incubated with 100 ng of glutathione-sepharose purified GST/LCRF-1 for 30 minutes at 4°C in 10 mM Tris pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, and 5% glycerol. Products were resolved on 4% nondenaturing polyacrylamide gels in 0.5×TBE. Oligonucleotides used for gelshift analysis were as follows:

AP1 5′ CAAGCACAGCAATGCTGAGTCATGATGAGTCATGCTGAGGC 3′
3′CGTGTCGTTACGACTCAGTACTACTCAGTACGACTCCGAAT 5′
G>T 5′ CAAGCACAGCAATTCTGAGTCATTATGAGTCATGCTGAGGC 3′
3′ CGTGTCGTTAAGACTCAGTAATACTCAGTACGACTCCGAAT 5′
A>C 5′ CAAGCACAGCAATGCTGCGTCATGATGCGTCATGCTGAGGC 3′
3′ CGTGTCGTTACGACGCAGTACTACGCAGTACGACTCCGAAT 5′
C/EBP 5′ AATTCAATTGGGCAATCAGG 3′
3′ GTTAACCCGTTAGTCCTTAA 5′.

Transactivation analysis

HS2 (AP1x2) γ/Luc, HS2 (AP1x2) tk/Luc, HS2 (Gal4x2) γ/Luc and HS2 (Gal4x2) tk/Luc reporter plasmids were constructed as follows. The 1.5 kb KpnI−BglII fragment of 5′ HS 2 was inserted into KpnI−BglII digested pGL2-Basic (Promega) to produce a plasmid designated HS2 (AP1x2) Luc. HS2 (Gal4x2) Luc was constructed by replacing the duplicated AP1-like sites in HS 2 with GAL4 binding sites using the megaprimer mutagenesis method (29). The outside primers overlapped the KpnI site and the BglII sites of the 1.5 kb 5′ HS 2 fragment. The mutagenesis primer was a 66 base oligonucleotide:

5′ CCTTCTGGCTCAAGCACAGCAATCCGGAGGACAGTCCTCCGGTGCTGAGGCTTA GGGTGTGTGCCC 3′.

HS2 (AP1x2) γ/Luc and HS2 (Gal4x2) γ/Luc were produced by digesting HS2 (AP1x2) Luc and HS2 (Gal4x2) Luc with BglII, blunting the ends with Klenow fragment of DNA polymerase and inserting a 335 bp AluI fragment containing human Aγ promoter and 5′ untranslated sequence from −299 to +36. HS2 (AP1x2) tk/Luc and HS2 (Gal4x2) tk/Luc were produced by digesting HS2 (AP1x2) Luc and HS2 (Gal4x2) Luc with BglII and inserting a 140 bp BglII PCR fragment containing HSV tk promoter and 5′ untranslated sequence from −120 to +20. The −120 to +20 tk promoter fragment with Bgl II ends was obtained by PCR amplification using the plasmid pREP4 (Invitrogen) as a template. The primers used were: −120, 5′ CCCGGGAGATCTATGT-CTTTAGTTCTATGATGACAC 3′; +20, 5′ CCCGGGAGA-TCTGCTGCAGGGTCGCTCGGTGTTCGA 3′. The PCR product was digested with BglII and subcloned into BglII digested HS2 (AP1x2) Luc and HS2 (Gal4x2) Luc.

The Gal4/LCR-F1 expression plasmid was constructed by inserting the 1.7 kb LCR-F1 cDNA fragment into the EcoRI site of pBXG1 (gift of V. Seyfert and M. Ptashne). This plasmid contains the SV40 promoter and enhancer inserted upstream of the Gal4 DNA binding domain (amino acids 1−147).

The Gal4/NF-E2 expression plasmid was constructed by inserting a 1.15 kb cDNA fragment into the EcoR1 site of pBXG1. The 1.15 kb cDNA clone was obtained by RT-PCR of Murine Erythroleukemia (MEL) cell polyA RNA using the cDNA Cycle Kit from Invitrogen. NF-E2 cDNA was amplified as three separate fragments which were subsequently ligated to produce the full length clone. The primers were based on the sequence of Andrews *et al.* (30) and are as follows. 5′GGATC-CGAATTCAGTAGGATGCCCCCGTGT 3′ and 5′TAAGGA-TACTCCACTGGG 3′were used to amplify an EcoRI−NspI fragment containing codons 1−175. 5′GGAGGAGCGAGTA-CGTGG 3′ and 5′CTGCCACCTTGTTCTTGC 3′ were used to amplify a NspI−SalI fragment containing codons 176 - 272. 5′C-CAGCTGGCTCTAGTTCG 3′ and 5′GGATCCGAATTCAG-TGTCCACTCTCTAGACCAG 3′ were used to amplify a SalI−EcoRI fragment containing codons 273−373 plus a stop codon. The full length clone was assembled in a four-way ligation

with EcoRI digested plasmid and the 1.15 kb cDNA was sequenced completely; no PCR artifacts were observed. The truncated clone was constructed by addition of an XbaI site immediately downstream of NspI in the EcoRI−NspI fragment and insertion of the resulting EcoRI−XbaI fragment into EcoRI−XbaI digested pBXG1.

Transfections were performed as follows. K562 cells were washed in serum-free RPMI 1640 media and 1.6 $\times 10^7$ cells in 500 μl of RPMI were added to a mixture of 3 μg reporter plasmid, 30 μg transactivator plasmid (1:10 molar ratio), and 10 μg pTKβ-gal (Clontech) internal control plasmid. The final volume was adjusted to 600 μl with Phosphate Buffered Saline (PBS). Electroporations were performed with a Bio-Rad Gene Pulser set at 320 V and 960 μF. After electroporation, the cells were diluted into 10 ml RPMI plus 10% Fetal Bovine Serum (FBS) in a 100 mM culture dish and incubated for 24 hours at 37°C in 5% CO_2. HeLa and HepG2 cells were transfected by the calcium phosphate co-precipitation method (26). Each 60 mm dish of cells (70% confluent) was transfected with the same constructs as described above except that an SV40/CAT plasmid (pCAT-Control; Promega) was used as an internal control. Twenty four hours after transfection, K562, Hela and HepG2 cells were harvested and extracts were prepared in Promega Reporter Lysis Buffer according to the manufacturer's instructions. Luciferase activity was determined by the Promega Luciferase Assay System, β-galactosidase was assayed as described by Miller (31) and CAT activity was determined by the CAT enzyme assay system (liquid scintillation modification) according the manufactures instructions (Promega). Luciferase activity was divided by β-galactosidase activity or CAT activity in each sample to control for different transfection efficiencies. Results from GAL4 + HS2 (GAL4x2) γ or tk/Luc co-transfections in K562, Hela or HepG2 cells were set at 1× and results of other transfections in the same cell type were normalized to this value. Experiments were performed at least three separate times in triplicate; the results are the mean of all experiments performed.

RESULTS AND DISCUSSION

Analysis of mutant HS 2 β constructs in transgenic mice

The G to T mutation discussed above was originally analyzed in constructs that were transfected into cultured erythroid cells. To confirm the functional significance of the G to T transversion, we analyzed the effects of this mutation on expression of a 5′ HS 2 β-globin construct in transgenic mice. Fertilized eggs were injected with a wild-type construct, a construct containing a deletion of the AP1-like sequences, a construct containing the G to T mutation or a control construct containing an A to C mutation as indicated in Figure 1. Embryos were removed at 16 days of gestation and fetal liver RNA was analyzed for correctly initiated human β-globin mRNA by primer extension (data not shown) and quantitated by solution hybridization (Figure 1). The wild-type HS 2 β construct was expressed at an average level of 42% of the endogenous mouse β-globin gene per gene copy and HS 2 ΔAP1 β was expressed at 2% of mouse β-globin per gene copy. These results are consistent with previous experiments in transgenic mice. The G to T mutations decreased the level of expression to 11% of mouse β-globin per gene copy. This diminution of HS 2 enhancer activity confirms that the G residues located 2 bp upstream of the consensus AP1 sites is required for efficient 5′ HS 2 activity. The A to C mutations did not affect expression.

Isolation of LCR-F1 cDNA clone

To isolate proteins that bind to the AP1-like sequences we constructed a λgt11 cDNA library with mRNA from K562 cells and screened this unamplified library with multimerized copies of the AP1-like sites (24, 25). The criteria for identifying specific clones were four-fold: (1) the fusion protein should bind to the

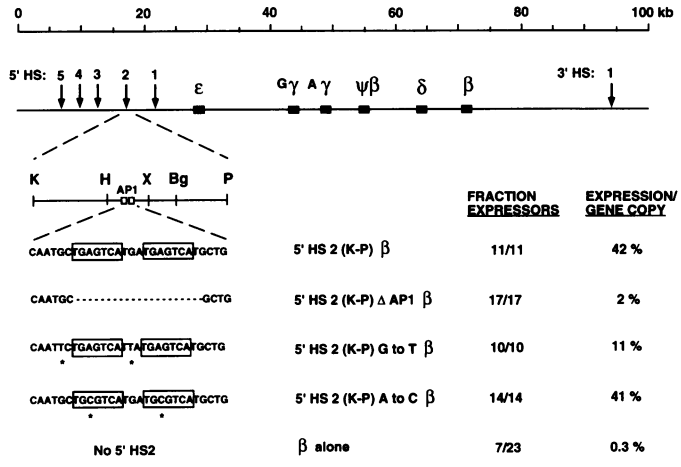

Figure 1. Expression of 5′ HS2 β-globin gene constructs in transgenic mice. The β-globin locus on human chromosome 11 is illustrated at the top. Black boxes represent individual globin genes, and arrows indicate DNase I hypersensitive sites that flank the locus. The 5′ HS 2 region is expanded to illustrate a 1.9 kb KpnI−PvuII fragment that was inserted upstream of a 4.1 kb HpaI−XbaI fragment containing the human β-globin gene. Restriction sites are as follows: K, KpnI; H, HindIII; X, XbaI; Bg, BglII; P, PvuII. Duplicated AP1-like sites are indicated by open boxes and an expanded view of this region is illustrated. Wild-type sequence and mutations within and surrounding the AP1-like sites are indicated below the 5′ HS 2 (K-P) fragment. Point mutations are marked by asterisks and the dashed line in the ΔAP1 construct represents an 18 bp deletion. The fraction of expressors represents the number of transgenic animals that express the transgene. Expression/gene copy is the mean percent expression per gene copy; this value is calculated as follows: (human β mRNA/human β gene copy)/(mouse β mRNA/ mouse β gene copy) ×100. The data for 5′ HS 2 (K-P) β and 5′ HS 2 (K-P) ΔAP1 b include data from Caterina *et al.* (21), and the data for the β alone construct are from Ryan *et al.* (9).

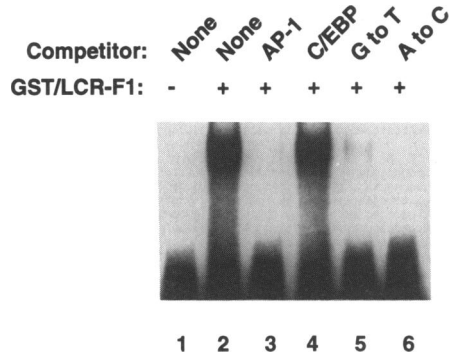

Figure 2. Gel shift analysis of LCR-F1. Purified GST/LCR-F1 fusion protein was incubated with an end-labeled oligonucleotide containing the duplicated AP1-like sites from 5′ HS 2 and a 100-fold molar excess of the indicated, unlabeled oligonucleotides. Protein−DNA complexes were separated from free oligonucleotides on 4% polyacrylamide gels.

multimerized AP1-like sites, (2) it should not bind to a multimerized C/EBP (CAAT/Enhancer Binding Protein) binding site, (3) it should not bind efficiently to multimerized AP1-like sites that contain the G to T mutation and (4) it should bind efficiently to multimerized AP1-like sites that contain the A to C mutation. The single clone that met these criteria was plaque-purified and designated LCR-F1. The 1.7 kb LCR-F1 insert was subcloned into pGSTag (32) and the binding specificity of purified GST/LCR-F1 fusion protein was determined by gel mobility shift analysis (27, 28). As indicated in Figure 2, GST/LCR-F1 bound specifically to a 41 bp oligonucleotide containing the duplicated AP1 sites (lane 2) and competition experiments (lanes 3–6) confirmed the binding specificity described above.

LCR-F1 is a novel bZIP protein that contains a CNC domain

The nucleotide sequence of LCR-F1 is illustrated in Figure 3A. The clone contains a nearly consensus Kozak sequence (33) that suggests an open reading frame of 447 amino acids. This

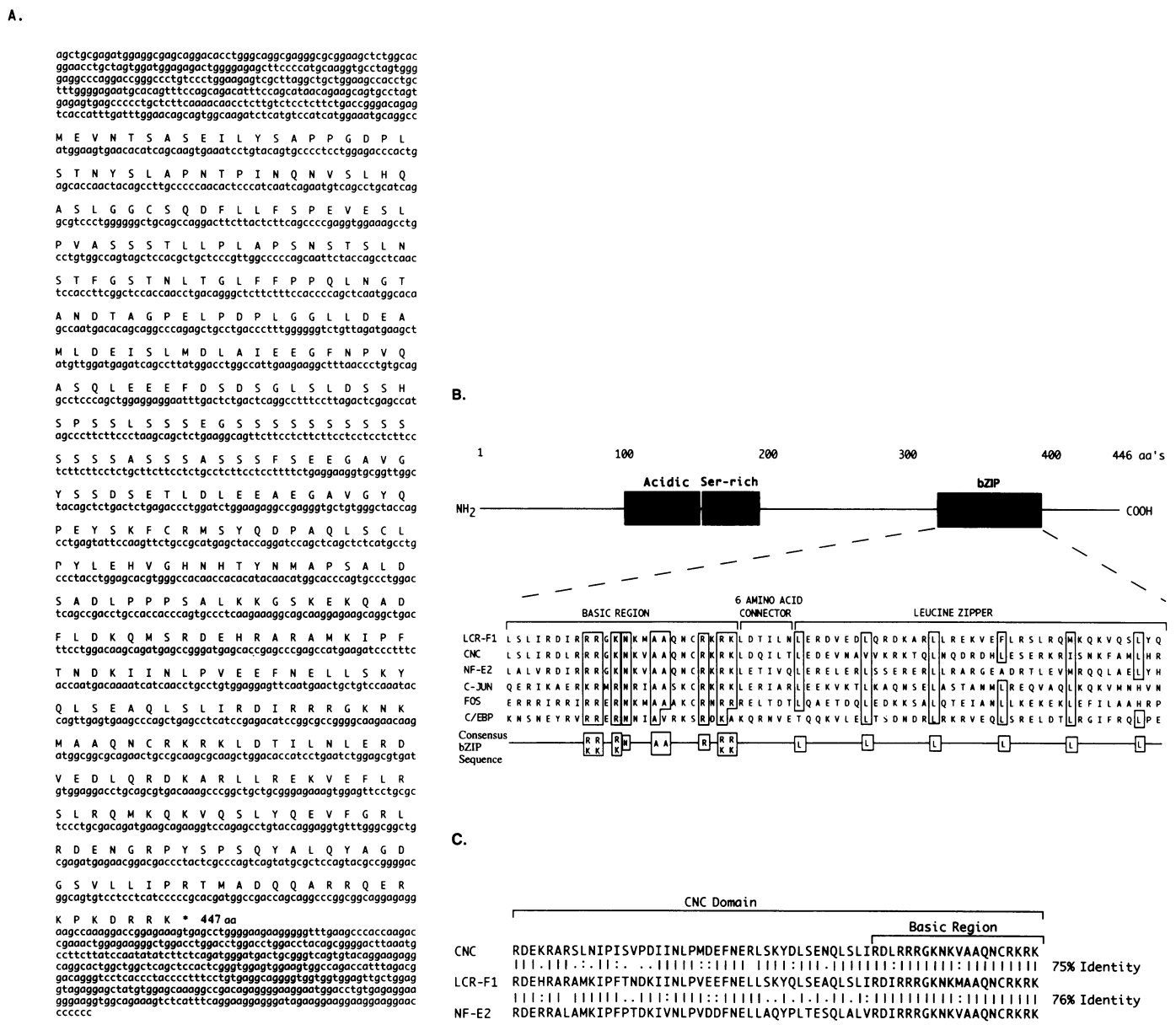

Figure 3. Sequence of LCR-F1. Panel **A.** Combined sequence of 3 overlapping LCR-F1 cDNA clones. The predicted amino acid sequence of LCR-F1 is listed above the DNA sequence. Panel **B.** Schematic of the LCR-F1 protein. The black boxes represent acidic, serine-rich and basic-leucine zipper domains of the protein. Below the schematic the bZIP domain of LCR-F1 is expanded and compared to 5 other bZIP proteins. Open boxes designate consensus bZIP amino acids. CNC is the Drosophila Cap N Collar protein (34) ; NF-E2 is a recently described erythroid-enriched factor that also binds to the 5′ HS 2 AP1-like sites (30) ; cJun, cFos and C/EBP homologies in this region have been described previously (50). Panel **C.** CNC domain. A 63 amino acid region of CNC, LCR-F1 and NF-E2 is compared. Lines represent identical amino acids; double dots represent conservative amino acid replacements; single dots represent semi-conservative amino acid replacements. The CNC domain overlaps the basic region of these 3 proteins. Outside the CNC domain LCR-F1, NF-E2 and CNC diverge completely except for limited homology in the leucine zipper region.

conclusion was confirmed by the detection of 3 in frame stop codons located 200 bp upstream of the Kozak sequence in an overlapping cDNA clone isolated from another K562 library. LCR-F1 has several striking features that are illustrated in Figure 3B. The carboxyl end of the protein contains a basic-leucine zipper (bZIP) domain. Three leucine residues spaced 7 amino acids apart are present at the amino terminal end of the zipper region. The fourth and fifth hydrophobic amino acids in the zipper are phenyalanine and methionine, and these are followed by 2 more leucine residues with correct spacing. Overall, the heptad repeats of the zipper region consist of 7 hydrophobic amino acids. A 6 amino acid linker separates the zipper and basic regions. The basic region contains lysine and arginine residues at positions identical to other bZIP proteins as well as the conserved asparagine located 10 amino acids upstream of the linker. Interestingly, LCR-F1 is 75% identical to the Drosophila Cap N Collar [CNC; (34)] protein and 76% identical to the erythroid-enriched factor NF-E2 (30) in a 63 amino acid domain that includes the basic region (Figure 3C). This extended basic region may be involved in DNA binding and, if so, is analogous to the POU domain located adjacent to POU-homeodomains in proteins such as Pit-1, Oct-1 and UNC 86 (35, 36). We propose to designate this region the 'CNC domain'. Presumably LCR-F1, NF-E2 and CNC are members of a larger family of proteins that contain this newly recognized motif (30).

Located approximately 135 amino acids amino-terminal to the CNC domain is a stretch of 40 amino acids that contains 30 serines (Fig 3). Within this region there is a patch of 14 serines in succession. Few proteins contain a domain that is as rich in serines. The serines may provide multiple sites for phosphorylation and/or glycosylation that modify the structure and/or activity of the protein. A 55 amino acid region adjacent to the serine-rich domain of the protein is highly acidic (Figure 3). This domain has a net charge of −14. As described below, the amino-terminal end of LCR-F1, which contains this acidic region, functions as a strong transactivation domain specifically in erythroid cells.

LCR-F1 is expressed in erythroid and non-erythroid cells

The Northern blot analysis illustrated in Figure 4A demonstrates that the LCR-F1 cDNA hybridizes to 5.0 kb and 4.5 kb mRNAs in K562, CEM-6 (T cell), NALM-6 (B cell), U937 (monocyte) and Hela cells. Murine erythroleukemia (MEL) cells contain 5.0 kb and 4.3 kb LCR-F1 mRNAs. Human bone marrow (BM) contains both mRNAs but, interestingly, only the 4.5 kb mRNA is observed in human reticulocytes (Figure 4B). We are presently isolating cDNA clones for both mRNAs to determine whether they encode distinct proteins.

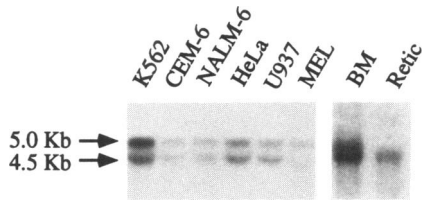

Figure 4. Northern blot analysis of LCR-F1 expression. Lanes 1−6 are 5 μg of total RNA from uninduced K562 (human erythroleukemia), CEM-6 (T cell), NALM-6 (B cell), Hela (cervical carcinoma), U937 (monocyte), uninduced MEL (murine erythroleukemia). Lanes 7 and 8 are 10 μg of total RNA from human bone marrow and human reticulocytes, respectively. The blots were hybridized with a randomly labeled 0.96 kb BamHI−EcoRI fragment containing the 3′ portion of the cDNA. A 5′ probe yielded identical results (data not shown).

Gal4/LCR-F1 fusion proteins activate high-level, erythroid-specific globin gene expression

Initial attempts to over-express LCR-F1 in K562 cells did not result in increased expression of a reporter construct composed of 5′ HS 2 sequences, a human γ-globin promoter, and a luciferase gene (HS2-γ/luc; Figure 5A). The level of HS 2-γ/luc expression in these experiments was high without exogenous LCR-F1, presumably due to saturating levels of endogenous LCR-F1. Background expression of the reporter gene was decreased by deleting the duplicated AP1 sites in 5′ HS 2 and replacing them in their natural context with duplicated consensus Gal4 binding sites (37) (Giniger *et al.* 1985; HS2(Gal4x2)-γ/luc, Figure 5). When the HS2(Gal4x2)-γ/luc construct was transfected into K562 cells, expression was 39-fold lower than the wild-type HS2(AP1x2)-γ/luc construct (Fig 5C). The LCR-F1 cDNA was then inserted in frame into an expression plasmid (pBXG1, gift of V. Seyfert and M. Ptashne) containing the Gal4 DNA binding domain (Gal4 1−147) and co-transfected with HS2(Gal4x2)-γ/luc into K562 cells. As illustrated in Figure 5C, the Gal4/LCR-F1 fusion protein enhanced HS2(Gal4x2)-γ/luc expression 62-fold. A mouse NF-E2 cDNA clone (30) was also inserted into pBXG1 and co-transfected with HS2(Gal4x2)-γ/luc into K562 cells. This clone also enhanced reporter gene expression (8-fold). As discussed below, both of these CNCbZIP proteins bind to the duplicated AP1 sites in HS2 and both may be involved in activating human globin gene expression.

Co-transformations of Gal4/LCR-F1 and HS2(Gal4x2)-γ/luc were also performed in Hela and HepG2 cells to determine whether transactivation was erythroid-specific. As indicated in Figure 5C, Gal4/LCR-F1 did not enhance expression in either cell type; reporter gene expression was identical with Gal4 and Gal4/LCR-F1 plasmids. Gal4/NF-E2 was also tested in Hela and HepG2 cells and like Gal4/LCR-F1 this protein failed to enhance expression. These results suggest that both LCR-F1 and NF-E2 transactivate reporter gene expression specifically in erythroid cells. Gal4/VP-16 was co-transfected with HS2(Gal4x2)-γ/luc into Hela and HepG2 cells as a control. Reporter expression was stimulated 5- and 8-fold, respectively; therefore, the γ-globin gene promoter can be activated in these non-erythroid cell types.

Gal4/LCRF1 was also assayed for the ability to transactivate a reporter gene containing a heterologous promoter. A HS2(Gal4x2)-tk/luc reporter gene was co-transfected with Gal4/LCRF1 or with Gal4/NF-E2 in K562, Hela and HepG2 cells (Figure 6). Gal4/LCRF1 transactivated this reporter gene 107-fold and Gal4/NF-E2 transactivated the reporter 7-fold in K562 cells. In Hela and HepG2 cells Gal4/LCRF1 activated the reporter only 3- and 12-fold, respectively; Gal4/NF-E2 activated the reporter only 1- and 3-fold, respectively. Gal4/VP-16 was co-transfected in Hela and HepG2 cells with HS2(Gal4x2)-tk/luc as a control, and reporter gene expression was stimulated 98- and 220-fold, respectively; this result demonstrates that the tk promoter can be activated at a high level in these non-erythroid cell types. The relatively low level of Gal4/LCRF1 and Gal4/NF-E2 transactivation in Hela and HepG2 cells confirms the hypothesis that efficient transactivation by both of these proteins is essentially erythroid-specific.

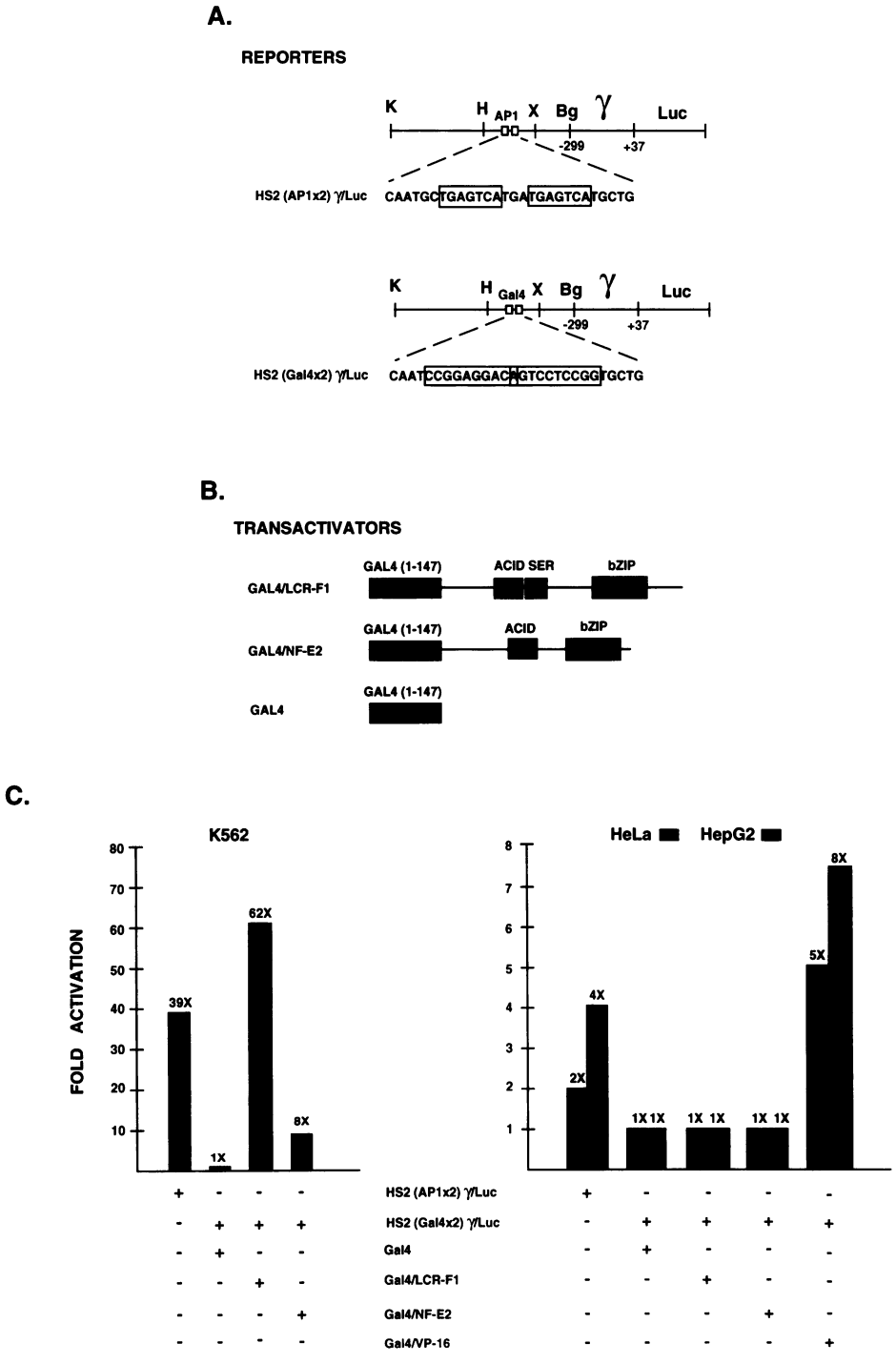

Figure 5. LCR-F1 activates high level, human globin gene expression specifically in erythroid cells. Panel **A.** Reporter constructs used in transfections. HS2 (AP1x2) γ/Luc contains wild-type 5′ HS 2 sequences on a 1.5 kb KpnI−BglII fragment (see Figure 1); this fragment is inserted upstream of a human $^{A}\gamma$ globin promoter that is linked to the luciferase gene. The duplicated AP1-like sites are boxed and an expanded view of this region is depicted below the fragment. HS2 (GAL4x2) γ/Luc contains the same 1.5 kb KpnI−BglII fragment as HS2 (AP1x2) γ/Luc except that the duplicated AP1-like sites have been replaced by duplicated GAL4 binding sites; the duplicated GAL4 binding sites, which overlap by one base, are boxed. Panel **B.** Transactivators used in transfection experiments. The black box designated GAL4 1−147 represents the 147 amino acid DNA binding domain of the yeast GAL4 protein. The GAL4 construct contains only the GAL4 DNA binding domain. GAL4/LCR-F1 is a fusion protein containing the GAL4 DNA binding domain fused to LCR-F1 and Gal4/NF-E2 is a fusion protein containing the GAL4 DNA binding domain fused to NF-E2. GAL4, GAL4/LCR-F1 and Gal4/NF-E2 expression are driven by an SV40 early promoter and enhancer. Panel **C.** Results of transfection experiments. The constructs listed below the graph and an internal control plasmid were transfected into K562, Hela and HepG2 cells and luciferase activity was determined as described in Materials and Methods.

LCR-F1 activation domain is an acidic region at the amino terminus

Figure 7 illustrates results of initial experiments to map the activation domain of LCR-F1. Deletion of the b/Zip region in the Gal4/LCR-F1 fusion protein increased expression of the reporter gene from 62-fold to 142-fold above expression of the reporter gene alone. Surprisingly, deletion of the serine rich region increased expression even further; reporter gene expression in this case was 177-fold above background. Although this result suggests that the serine domain is not required for activity of the fusion protein, the serines may play an important structural role in the native protein. Serine modification may alter the structure of the native protein so that the powerful activation domain can be positioned in a specific orientation to maximally stimulate expression.

Deletion of the carboxyl terminal sequences of NF-E2 also stimulated expression; this truncated protein enhanced expression of HS2(Gal4x2)-γ/luc 18-fold. A stretch of 52 amino acids in this truncated protein contains a net negative charge of −12. Therefore, both LCR-F1 and NF-E2 contain strong acidic activating domains.

Gal4/VP-16 was also co-transfected into K562 cells with HS2(Gal4x2)-γ/luc for comparison with Gal4/LCR-F1 and

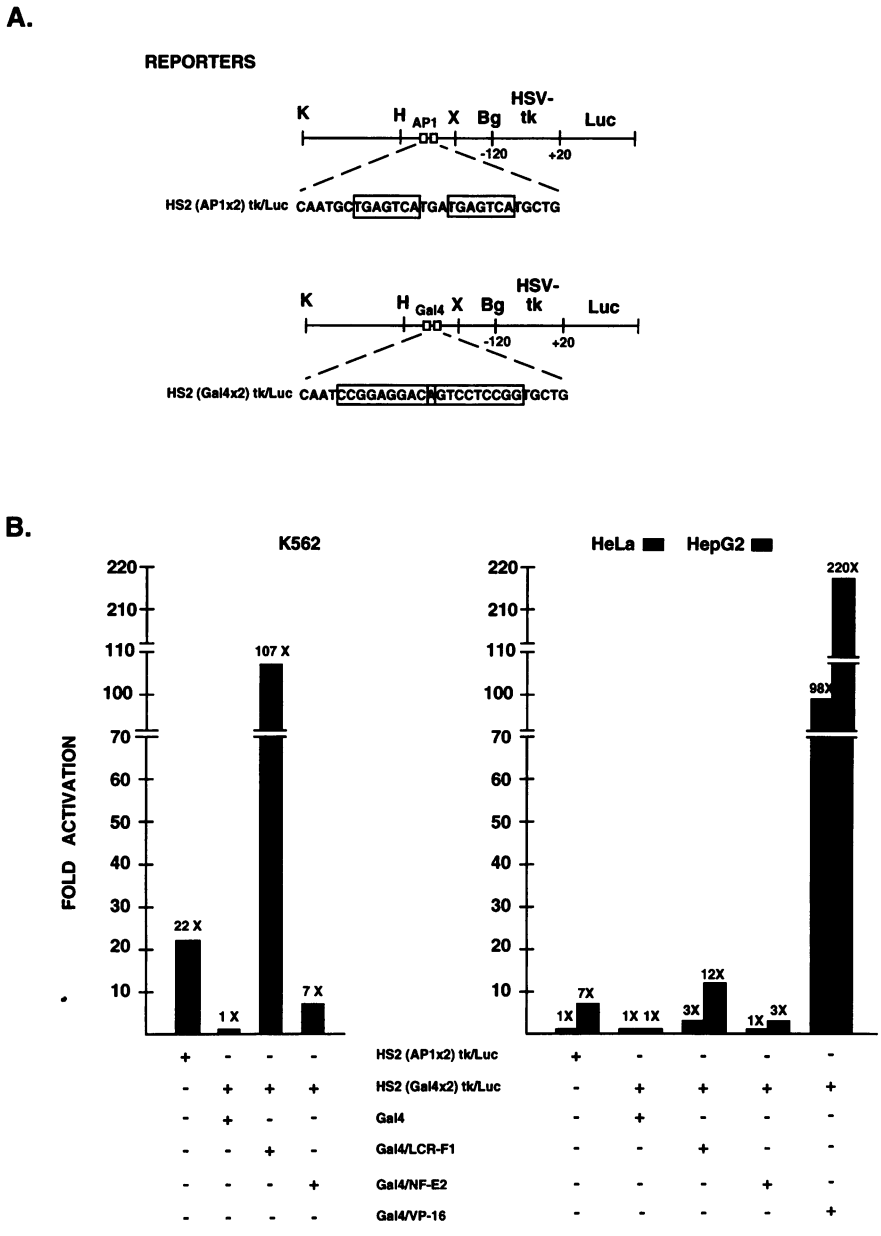

Figure 6. LCR-F1 transactivates a heterologous promoter specifically in erythroid cells. Panel **A.** Reporter constructs used in transfections. The reporters were identical to the reporters illustrated in Figure 5 except that the tk promoter and 5′ untranslated sequence (−120 to +20) replaced the γ-globin promoter and 5′ untranslated sequence (−299 to +37). Panel **B.** Results of transfection experiments. The constructs listed below the graph and an internal control plasmid were transfected into K562, Hela and HepG2 cells and luciferase activity was determined as described in Materials and Methods.

Gal4/NF-E2. The Gal4/VP-16 fusion protein transactivated the reporter 167-fold. This is approximately the same level of activation observed with Gal4/LCR-F1 and is somewhat higher than the activity observed with GAL4/NF-E2.

Relationship between LCR-F1 and NF-E2

Andrews *et al.* (30) recently described the cloning of NF-E2 from murine erythroleukemia (MEL) cells and Ney *et al.* (38) subsequently reported the cloning and characterization of human NF-E2. This protein is part of a heterodimeric complex which also binds to the duplicated AP1 sites in 5′ HS 2 *in vitro*. As discussed above, NF-E2 and LCR-F1 are both bZIP proteins that are 76% identical in a 63 amino acid CNC domain which includes the basic region. Although the 2 proteins are highly homologous in this region, they diverge outside the CNC domain except for limited homology in the leucine zipper.

What is the relationship between NF-E2 and LCR-F1 in erythroid cells? The proteins may form heterodimers in vivo. However, immunoprecipitations of erythroid nuclear extracts with antiserum against NF-E2 reveal only NF-E2 (45 Kd) and a smaller protein (18 Kd) (30, 38, 39) Alternatively, LCR-F1 and NF-E2 may function at different developmental stages; 5′ HS 2 activates globin gene expression in embryonic, fetal and adult life (13, 40–43) and different protein complexes that bind to the AP1-like sites may be involved in expression at these different stages.

LCR-F1 and erythroid-specific globin gene expression

Although LCR-F1 is not erythroid-specific, efficient transactivation of human globin gene expression by this protein is observed specifically in erythroid cells. There are several possible explanations for this observation. LCR-F1 may enhance expression only after interacting with the erythroid-enriched protein GATA-1 (44, 45) that binds approximately 50 bp downstream of the AP1-like sites in 5′ HS 2 (19). Alternatively, LCR-F1 may enhance expression through an erythroid-specific co-activator. An interesting precedence for this latter model has recently been described. An interaction between the ubiquitous transcription factor Oct-1 and a B lymphocyte specific co-activator may be responsible for B-cell specific immunoglobulin gene expression. This conclusion is supported by 2 pieces of data. First, Oct-1 plus a purified B-cell specific factor OCT-B efficiently transcribe immunoglobulin genes *in vitro* (46) ; secondly, immunoglobulin expression is not inhibited in mice (47) or in cultured B cells (48) by a knock out mutation in the Oct 2 gene. Although Oct 2 is B-cell-specific, the protein is not required for B-cell-specific immunoglobulin gene expression.

After this work was completed, Chan and Kan (49) published the sequence of Nrf1 (NF-E2 related factor 1) which is highly homologous to LCR-F1. LCR-F1 and Nrf1 cDNAs are identical except that LCRF1 contains a 91 bp insertion located 119 bp upstream of the predicted translational start site. This insertion, which is most likely an untranslated exon, produces an in frame stop codon 241 bp upstream of the translational start site. The LCR-F1 open reading frame is 447 codons whereas Nrf1 is 760 codons. Southern blot hybridizations suggest that a single gene encodes LCR-F1 and Nrf1.

In summary, we have cloned and characterized a novel human transcription factor (LCR-F1) that binds to β-globin LCR sequences and strongly activates human globin gene expression. LCR-F1 contains a 63 amino acid segment that has been designated the CNC domain; this domain is greater than 70%

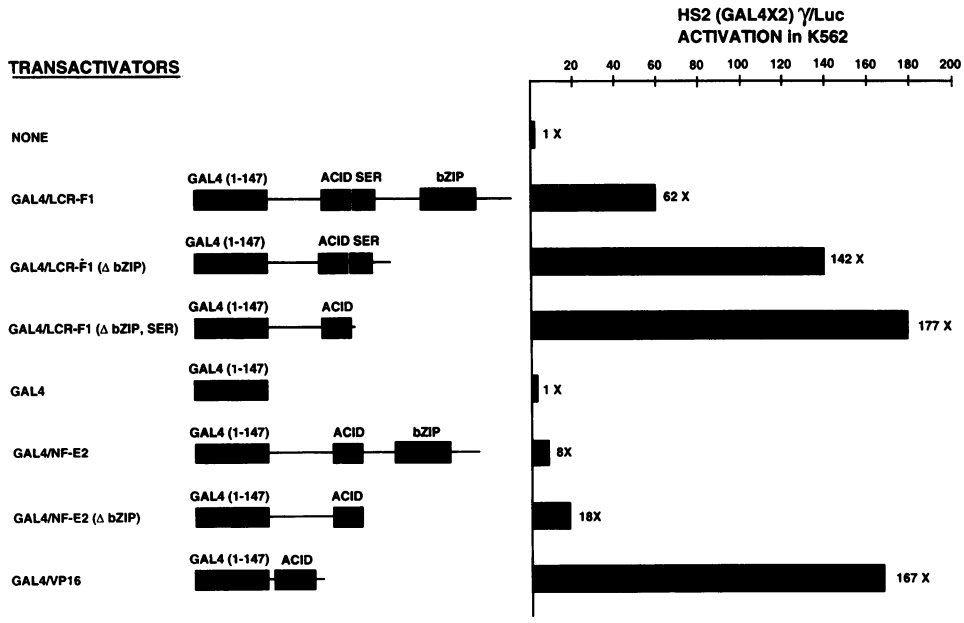

Figure 7. Mapping the LCR-F1 transactivation domain. The GAL4, GAL4/LCR-F1 and GAL4/NF-E2 transactivators and HS2(GAL4x2)γ/Luc reporter were as described in Figure 5. GAL4/LCR-F1 (ΔbZIP) encodes a fusion protein with a truncation of the carboxyl-terminal 214 amino acids; this deletes the bZIP domain. GAL4/LCR-F1 (ΔbZIP, SER) encodes a fusion protein with a truncation of the carboxyl-terminal 289 amino acids; this mutation deletes the bZIP and serine-rich domains. GAL4/NF-E2 ΔbZIP encodes a fusion protein with a truncation of the carboxyl-terminal 198 amino acids; this mutation deletes the bZIP region. The transactivator clones, which were driven by the SV40 promoter, the HS2 (GAL4x2) γ/Luc reporter and a tk/lacZ plamid were co-transfected into K562 and activity was determined as described in the legend to Figure 5.

identical in the Drosophila segmentation protein Cap N Collar (34) , the erythroid-enriched protein NF-E2 (30) and LCR-F1. LCR-F1 is not erythroid-specific but activates globin gene expression preferentially in erythroid cells. These results suggest that LCR-F1 may interact with other proteins, perhaps erythroid-specific co-activators, to direct high-level expression of human ϵ-, γ- and β-globin genes.

ACKNOWLEDGEMENTS

We especially thank Jill Lahti who was invovled in the initial stages of this work. We also thank Dr. Carl Pinkert and entire staff of the UAB Transgenic Animal Facility for producing the transgenic animals described in this study. We are grateful to Vicki Seyfert and Mark Ptashne for providing the pBXG1 and pBXGAL-VP16 plasmids. This work was supported in part by grants HL35559 and HL43508 from the National Institutes of Health. Support for synthesis of oligonucleotides for DNA sequencing and site-directed mutagenesis was provided through NCI grant CA13148 to the Comprehensive Cancer Center. Support for DNA sequence analysis computer prograns used in this work was provided by NIH Centers for AIDS Research program grant P30 AI27767. John Caterina is a graduate fellow supported by National Institutes of Health grant 5 T32 CA09467. David Donze is a graduate fellow supported by National Institutes of Health grant T32 GM08111.

REFERENCES

1. Tuan, D., Solomon, W., Li, Q. and London, I.M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6384–6388.
2. Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 1359–1363.
3. Forrester, W.C., Takegawa, S., Papayannopoulou, T., Stamatoyannopoulos, G. and Groudine, M. (1987) *Nucleic Acids Res.*, **15**, 10159–10177.
4. Orkin, S.H. (1990) *Cell*, **63**, 665–672.
5. Townes, T.M. and Behringer, R.R. (1990) *Trends Genet.*, **6**, 219–223.
6. Dillon, N. and Grosveld, F. (1993) *Trends Genet.*, **9**, 134–137.
7. Engel, J.D. (1993) *Trends Genet.*, **9**, 304–309.
8. Grosveld, F., van, A.G., Greaves, D.R. and Kollias, G. (1987) *Cell*, **51**, 975–985.
9. Ryan, T.M., Behringer, R.R., Martin, N.C., Townes, T.M., Palmiter, R.D. and Brinster, R.L. (1989) *Genes & Dev.*, **3**, 314–323.
10. Curtin, P.T., Liu, D., Liu, W., Chang, J.C. and Kan, Y.W. (1989) *Proc. Natl. Acad. Sci., USA*, **86**, 7082–7086.
11. Collis, P., Antoniou, M. and Grosveld, F. (1990) *EMBO J*, **9**, 233–240.
12. Fraser, P., Hurst, J., Collis, P. and Grosveld, F. (1990) *Nucleic Acids Res.*, **18**, 3503–3508.
13. Fraser, P., Pruzina, S., Antoniou, M. and Grosveld, F. (1993) *Genes & Dev.*, **7**, 106–113.
14. Chung, J.H., Whiteley, M. and Felsenfeld, G. (1993) *Cell*, **74**, 505–514.
15. Driscoll, M.C., Dobkin, C.S. and Alter, B.P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7470–7474.
16. Forrester, W.C., Epner, E., Enver, T., Driscoll, M.C., Papayannopoulou, T. and Groudine, M. (1990) *Genes & Dev.*, **4**, 1637–1649.
17. Ney, P.A., Sorrentino, B.P., McDonagh, K.T. and Nienhuis, A.W. (1990) *Genes & Dev.*, **4**, 993–1006.
18. Moi, P. and Kan, Y.W. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 9000–9004.
19. Talbot, D., Philipsen, S., Fraser, P. and Grosveld, F. (1990) *EMBO J.*, **9**, 2169–2178.
20. Liu, D., Chang, J.C., Moi, P., Liu, W., Kan, Y.W. and Curtin, P.T. (1992) *Proc. Natl. Acad. Sci. U.S.A.*, **89**, 3899–3903.
21. Caterina, J.J., Ryan, T.M., Pawlik, K.M., Palmiter, R.D., Brinster, R.L., Behringer, R.R. and Townes, T.M. (1991) *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 1626–1630.
22. Talbot, D. and Grosveld, F. (1991) *EMBO J.*, **10**, 1391–1398.
23. Townes, T.M., Lingrel, J.B., Chen, H.Y., Brinster, R.L. and Palmiter, R.D. (1985) *EMBO J.*, **4**, 1715–1723.
24. Singh, H., LeBowitz, J.H., Baldwin, A.S. and Sharp, P.A. (1988) *Cell*, **52**, 415–423.
25. Vinson, C.R., LaMarco, K.L., Johnson, P.F., Landschulz, W.H. and McKnight, S.L. (1988) *Genes & Dev.*, **2**, 801–806.
26. Sambrook, J., Fritch, E.F. and Maniatis, T. (1989) Molecular Cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
27. Garner, M.M. and Revzin, A. (1981) *Nucleic Acids Res.*, **9**, 3047–3060.
28. Fried, M. and Crothers, D.M. (1981) *Nucleic Acids Res.*, **9**, 6505–6525.
29. Sarkar, G. and Sommer, S.S. (1990) *BioTechniques*, **8**, 404–407.
30. Andrews, N.C., Erdjument-Bromage, H., Davidson, M.B., Tempst, P. and Orkin, S.H. (1993) *Nature*, **362**, 722–728.
31. Miller, J. (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
32. Ron, D. and Dressler, H. (1992) *BioTechniques*, **13**, 866–869.
33. Kozak, M. (1984) *Nucleic Acids Res.*, **12**, 857.
34. Mohler, J., Vani, K., Leung, S. and Epstein, A. (1991) *Mech. Dev.*, **34**, 3–10.
35. Herr, W., Sturm, R.A., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) *Genes & Dev.*, **2**, 1513–1516.
36. Rosenfeld, M. (1991) *Genes & Dev.*, **5**, 897–907.
37. Giniger, E., Varnum, S. and Ptashne, M. (1985) *Cell*, **40**, 767–774.
38. Ney, P.A., Andrews, N.C., Jane, S.M., Safer, B., Purucker, M.E., Weremowicz, S., Morton, C.C., Goff, S.C., Orkin, S.H. and Nienhuis, A.W. (1993) *Mol. Cell. Biol.*, **13**, 5604–5612.
39. Andrews, N.C., Kotkow, K.J., Ney, P.A., Erdjument-Bromage, H., Tempst, P. and Orkin, S.H. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 11488–11492.
40. Behringer, R.R., Ryan, T.M., Palmiter, R.D., Brinster, R.L. and Townes, T.M. (1990) *Genes & Dev.*, **4**, 380–389.
41. Behringer, R.R., Ryan, T.M., Reilly, M.P., Asakura, T., Palmiter, R.D., Brinster, R.L. and Townes, T.M. (1989) *Science*, **245**, 971–973.
42. Lloyd, J., Krakowsky, J., Crable, S. and Lingrel, J. (1992) *Mol. Cell. Biol.*, **12**, 1561–1567.
43. Morley, B.J., Abbott, C.A., Sharpe, J.A., Lida, J., Chan, T.P.S. and Wood, W.G. (1992) *Mol. Cell. Biol.*, **12**, 2057–2066.
44. Tsai, S.-F., Martin, D.I.K., Zon, L.I., D'Andrea, A.D., Wong, G.G. and Orkin, S.H. (1989) *Nature*, **339**, 446–451.
45. Evans, T. and Felsenfeld, G. (1989) *Cell*, **58**, 877–885.
46. Luo, Y., Fujii, H., Gerster, T. and Roeder, R. (1992) *Cell*, **71**, 231–241.
47. Corcoran, L.M., Karvelas, M., Nossal, G.J.V., Ye, Z.-S., Jacks, T. and Baltimore, D. (1993) *Genes & Dev.*, **7**, 570–582.
48. Feldhaus, A.L., Klug, C.A., Arvin, K.L. and Singh, H. (1993) *EMBO J.*, **12**, 2763–2772.
49. Chan, J.Y., Han, X.-L. and Kan, Y.W. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 11371–11375.
50. Landschulz, W.H., Johnson, P.F. and McKnight, S.L. (1988) *Science*, **240**, 1759–1964.